Claims

635   What is claimed is:

1. A method for collecting global or population characteristics for decision tree regulation comprises the following steps:
    (a) Input a decision tree;

640
    (b) Input a set of training samples;
    (c) Use the training samples to determine a decision characteristic for at least one decision tree node, said decision characteristic selected from the group consisting of global characteristics and population characteristics.

645
2. The method of claim 1 wherein the decision characteristic compensates for unequal class prevalence in the training samples.

3. The method of claim 1 wherein the decision characteristic compensates for errors in the training data.

650
4. The method of claim 1 wherein the global characteristics include global counts.

5. The method of claim 1 wherein the global characteristics include global population statistic.

655
6. The method of claim 1 wherein the population characteristics include local population statistic.

7. A method for classification regulation by information integration comprises the
660   following steps:
    (a) Input a decision tree;

(b) Input a plurality of decision characteristics selected from the group consisting of global characteristics and population characteristics from at least one terminal node of the decision tree;   .

665

(c) Determine the confidence value for each of the plurality of said decision characteristics

(d) Determine an integrated confidence value for each class of said at least one terminal node.

670

8. For a crisp tree application, the method of claim 7 further assigns the class with the maximum integrated confidence value as the decision for the terminal node.

9. For a smooth tree application the method of claim 7 further uses the integrated confidence value as the likelihood value.

675

10. The method of claim 7 wherein the global characteristics and population characteristics are selected from the group consisting of global counts, local counts, global population statistic, and local population statistic.

680

11. The method of claim 7 wherein the confidence value is selected from the set consisting of local count confidence, local population confidence, global count confidence and global population confidence.

12. The method of claim 7 wherein the integrated confidence value is a weighted

685

combination of a plurality of confidence values.

13. The method of claim 7 wherein the global characteristics have a global context coverage that is adjusted using different layer depths.

690

14. The method of claim 7 wherein the global characteristics have a global context coverage that is adjusted on a minimum number of training samples.

15. A method for tree pruning regulation by information integration comprises the following steps:

695   (a) Input a decision tree;

    (b) Input a set of training samples;

    (c) Generate a regulated measure selected from the group consisting of integrated confidence values and reliability measures;

    (d) For a non-terminal node of the tree having two descending terminal nodes,

700   determine the accuracy values using the regulated measure under two separate nodes or combined node conditions;

    (e) If combined node accuracy value is greater than the two separate node accuracy value, prune the terminal nodes by combing the two terminal nodes and convert the associated non-terminal nodes into one terminal node.

705

16. The method of claim 15 wherein the reliability measures include a local population reliability measure.

17. The method of claim 15 wherein the reliability measures include a count
710   reliability measure.

18. The method of claim 15 wherein the reliability measures include a population reliability measure.

715   19. The method of claim 15 wherein the reliability measures include a combined reliability measure.

20. The method of claim 15 wherein the reliability measures include a global population reliability measure.

720

21. The method of claim 15 wherein the reliability measures include a combined reliability measure.

22. The method of claim 15 wherein the reliability measure for the maximum class is
integrated with the classification accuracy as the criteria for tree pruning

23. A method for tree generation regulation by information integration comprises the
following steps
    (a) Input a set of training samples;
    (b) For at least one node, generate a set of candidate thresholds;
    (c) Partition data at a candidate threshold;
    (d) Calculate an evaluation function selected from the set consisting of
        integrated confidence value and reliability measures;
    (e) Select the partition for the node as the one that maximizes the
        evaluation function.

725

730

735

740